



**QUEEN'S
UNIVERSITY
BELFAST**

Selection Power and Selection Labor for Information Retrieval

Warner, J. (2007). Selection Power and Selection Labor for Information Retrieval. *Journal of the American Society for Information Science and Technology*, 58(7), 915-923. <https://doi.org/10.1002/asi.20576>

Published in:

Journal of the American Society for Information Science and Technology

Document Version:

Early version, also known as pre-print

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright © 2007 Wiley Periodicals, Inc., A Wiley Company.

This is the pre-peer reviewed version of this article, which has been published in final form at doi: 10.1002/asi.20576. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

This is a preprint of a paper published by the Journal of the American Society for Information Science and Technology © 2007 Wiley. The published paper can be retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/asi.20576/abstract>.

Selection power and selection labor for information retrieval

Julian Warner

**School of Management and Economics
The Queen's University of Belfast**

j.warner@qub.ac.uk

Abstract. This paper is concerned with the relation between selection power and selection labor for information retrieval. It is the first part of the development of a labor theoretic approach to information retrieval.

Existing models for evaluation of information retrieval systems are reviewed and the distinction of operational from experimental systems partly dissolved. The often covert, but powerful, influence from technology on practice and theory is rendered explicit.

Selection power is understood as the human ability to make informed choices between objects or representations of objects and is adopted as the primary value for information retrieval. Selection power is conceived as a property of human consciousness, which can be assisted or frustrated by system design. The concept of selection power is further elucidated, and its value supported, by an example of the discrimination enabled by index descriptions, the discovery of analogous concepts in partly independent scholarly and wider public discourses, and its embodiment in the design and use of systems.

Selection power is regarded as produced by selection labor, with the nature of that labor changing with different historical conditions and concurrent information technologies. Selection labor can itself be decomposed into description and search labor.

Selection labor, and its decomposition into description and search labor, will be treated in a subsequent paper, in a further development of a labor theoretic approach to information retrieval.

Introduction

Labor, choice, and technology are fundamental to human experience. Once out of Eden, we are condemned to labor and compelled to choose, in the Judeo-Christian tradition. Technology may have been less explicitly noticed but has been equally pervasive in post-Edenic experience, both as productive and information technologies. Physical and mental labor have tended to be considered separately from each other. There have been more recent moves towards synthesis in the acknowledgment of mental components in physical labor and physical elements in mental labor, with an emerging view of intelligence as a 'quality of our bodies as much as our minds' (Gosden, 2003, pp.31-33, 119). Mental or informational labor has also been recognized as an activity in itself, as well as an adjunct to obtaining physical control over the environment (Webster, 2002, p.15). Types of mental labor have been differentiated, with semantic labor distinguished

from syntactic labor (Warner, 2005a). Choice can be regarded as the product of mental labor, classically, from Aristotle, as deliberation. Late 20th century developments in information technology have been viewed as constituting a revolution in the mechanization of mental labor (Minsky, 1967, p.2), embodied in the computer as a universal information machine, developing from mid- and late-19th century antecedents in special purpose information machines. Both productive and information technology are human constructions, specifically the products of human physical and intellectual labor upon natural resources and pre-existing human constructions (Warner, 2004, pp.5-35). An understanding of, and evaluative model for, information retrieval constructed from labor, choice, and technology promises to be deeply rooted in human experience and to offer a radical depth of understanding.

The model proposed must have a dual aspect, comprehending empirical reality and selectively absorbing existing models, if it is to be regarded as a scientific advance. Empirical reality should be explained as fully, powerfully, and as parsimoniously as possible. A strong degree of correspondence to empirical reality is promised by the pervasive presence of labor, choice, and technology in information retrieval practice. Human labor is immediately present as the description labor of cataloging, classification, and database description. Choice has been persistently embodied in practice, and, more recently, increasingly theoretically recognized and valued as selection from retrieval results and on filtering of information. Modern information technologies, to which aspects of human mental labor can be, and increasingly are, transferred, diffused from the 1950s and are now pervasive in information retrieval.

The overall intention is to develop a labor theoretic approach to information retrieval, and the immediate concern, in this article, is with the initial components for that approach. Existing evaluative traditions will be reviewed and the possibility for synthesis within a labor theoretic approach indicated. Selection power and selection labor will then be introduced and developed, both as concepts and activities in themselves and for the relation between them.

Existing models

Power in explanation can be demonstrated by the ability to absorb elements from previous models as special cases of the new model, indicative of the history of a true science, while discarding those elements which have obstructed understanding. Existing evaluative models which should be absorbed into the new model, and which do offer some elements for synthesis and carrying forward, have been developed directly in information science, with diffusion into computer science, and in librarianship and indexing. Information society concerns with informational or mental labor offer a more indirect resource which can also be absorbed.

The dominant tradition for the evaluation of information retrieval systems emerged, nearly simultaneously and partly independently, in both the United States and the United Kingdom, in information science in the early to mid-1950s (Ellis, 1996, pp.1-22), and has since diffused to, and been partly absorbed in, computer science. The techniques

developed for selection and ordering of references and documents have served as exemplars, or demonstrations of possibility, for the increasingly dominant Internet search engines, with some elements of more direct transfer or inheritance in search algorithms derived from information retrieval research (Ellis and Vasconcelos, 1999, p.8). There were also parallel, largely conceptually independent, but drawing on common technologies, developments in commercial search services, similarly serving as exemplars and demonstrating commercial feasibility as well as technical possibility. The information and computer science tradition has not always been explicit about its own values or examined its assumptions and has sometimes pragmatically, and without full notice, departed from its initial assumptions. It can, however, be broadly characterized as query transformation, with the query verbally articulated in advance of searching, and then transformed by a system into a set of records (Heine, 1980). Retrieved records are then assessed for their relevance to the query and measures of performance, including precision and recall, are generated. The methodology adopted induces a bias towards fixing and possibly reifying relevance, reducing it from a concept to a relation between a query and a document (Ellis, 1984, pp.28-29; 1992; 1996, pp.11-20). Bibliographic, rather than full text, systems have been the dominant, although not exclusive, subject of study, with humanly assigned indexing received as a given and the rationale for such indexes, and the labor and costs involved in their making, not explored. There has also been an implicit teleology, with a perfect system as a goal, and evaluation has been raised to an end in itself, sometimes obstructing understanding (Ellis, 1984; 1992).

The relation of the information and computer science research tradition to information technology could be characterized as repression, with the repressed reemerging but not fully consciously known. Repression is discernible in the insistence that the retrieval processes created and studied are independent of their particular technological instantiation, while, simultaneously, allowing procedures to be strongly determined by contemporary technological possibilities. For instance, the stress on query transformation corresponded to the batch-processing embodied in the technology of the 1950s. The theoretical legacy of query transformation has proved difficult to adapt to modern systems, which do not necessarily demand a query verbally articulated in advance of searching and which can be, and are, used interactively. Critiques argued that the assumption of the necessity for a verbally articulated query was intra-theoretic (Heine, 1980), rather than intrinsic to information seeking, and this has been substantiated by subsequent technological developments. The reemergence of the repressed can be found in the late articulation, and still limited acknowledgement of the identity between the primitive operations of information retrieval and of logic or computation. The potential transformations for information retrieval on written records or descriptions have been revealed on analysis to be variations on primitive operations of sorting or partitioning and the transformation of one symbol into another (Buckland and Plaunt, 1994). This can be regarded as a special case of the known potential for reducing mathematical and logical operations on an object-language to the writing, erasure and substitution of symbols (Ramsey, 1990, pp.165-174), also corresponding to the primitive computational operations (Warner, 1994, pp.102-103). The paradigm of query transformation can be regarded as largely, although not entirely, exhausted, becoming increasingly distant from the empirical reality of interactive and distributed

(Ellis and Vasconcelos, 1999, p.8) systems, exposing its rigidity if the original distinctions are retained, or surviving by *ad hoc* modifications to its theoretical base, losing relevance in the first direction and internal intellectual coherence in the other.

Technology can be fully acknowledged rather than repressed, and a distinction between techniques and values introduced and in order to preserve and carry forward what may be valuable from information retrieval research in information and computer science. The utility of techniques developed for selection and ordering of references and documents, in both the experimental tradition and commercial practice, can be acknowledged, while simultaneously recognizing that these techniques are derived from known fundamental computational operations, realized in the special purpose tools and machines used for information retrieval at the beginning of the research tradition in the 1950s and by the programmed universal information machine of the modern computer. Values for information retrieval can be brought into accord with processes by replacing query transformation by selection power. A transfer of human mental labor to technology, and the dynamic of relative costs which has effectively compelled that transfer, has been identified (Warner, 2005b) and can be specifically adapted to an understanding of information retrieval developments. Constraints have been established for the types of mental labor transferable to technology, from the distinction between semantic and syntactic labor (Warner, 2005a), and can be further developed in particular relation to information retrieval.

Librarianship and indexing have longer historical antecedents for understanding information retrieval than the research tradition developed in information science and subsequently diffused to computer science, but are currently less widely influential. They have tended to be less explicit about their evaluative criteria and aims for information retrieval systems and far less concerned with producing measures of effectiveness. In contrast to information and computer science, they have been associated with the technologies of writing and of printing, rather than the computer, and have had a pronounced preference for direct human description of information objects. A similar pattern of repression, although less immediately pronounced, can be discerned with regard to technology: the need for descriptions more concise than the documents described, imposed by the storage constraints of media for inscription, and for direct human intervention in the creation of these descriptions, connected with the technical characteristics of writing and printing, have tended to be universalized and treated as they were independent of their dominant technological realizations (Wilson, 2001). The assumption of the need for index descriptions more concise than the documents described was inherited by the information and computer science tradition, although probably directly from existing information products rather than from the theories which informed the construction of those products (Cleverdon, 1962; Cleverdon, Mills, and Keen, 1966). A further limitation of library studies has been a focus on training in the use of information retrieval systems, often concentrating on the level of system commands, rather than understanding of their value in communication (Roberts, 1989). Where the role of formal information retrieval systems in communication, particularly scholarly communication, has been studied, there is disturbing evidence of their marginality, particularly in the sense of information, topic, or subject retrieval, rather than document

identification, by known author or title, and supply (Bath University Library, 1971; Smithson, 1994). The most valuable elements to be carried forward from librarianship and indexing are: a partly implicit stress on selection power, conceived as bibliographic control in librarianship (Wilson, 1968) and implied by the valuing of the discriminatory power of index terms in indexing, but made fully explicit here; and an acknowledgement of the role of direct human intellectual labor in creating selection power, transformed into a fuller understanding distinguished from specific technological constraints and their, partly covert, influence on theory and practice.

Information society discussions have given some, although rather limited, attention to information retrieval. For instance, Lyotard comments:

It is reasonable to suppose that the proliferation of information-processing machines is having, and will continue to have, as much an effect on the circulation of learning as did advancements in human circulation (transportation systems) and later, in the circulation of sounds and visual images (the media).

(Lyotard, 1984, p.4)

Other comments remain similarly rather unfocused, recognizing the significance of information retrieval, but not providing a full research or intellectual context for its consideration. In particular, technology has been unsatisfactorily treated in some information society discussions (Webster, 2002), possibly from wariness of being stigmatized as technologically determinist (Wilson, 1996a), and there has been a limited understanding of fundamental computer operations. An analytically valuable category of informational labor has begun to be distinguished (Webster, 2002, p.15) and this will be adopted and further differentiated, with the possibility of the transfer of some forms of mental labor to information technology acknowledged.

Different elements from the information retrieval tradition developed in information science, from librarianship and indexing, and from information society discussions will, then, be selected and carried forward. The utility of the techniques developed by information retrieval research, but not the associated value of query transformation, is acknowledged, while recalling that techniques are variations on primitive computational transformations. Selection power is adopted from librarianship and indexing as the primary value and the role of direct human labor is both substantiated and critiqued. Informational labor, transformed into mental labor to incorporate its historical antecedents, is taken from information society discussions. Technology is restored not repressed, and the understanding of the types of mental labor transferable to information technology is informed by the distinction between semantic and syntactic mental labor. A synthesis of existing approaches is envisaged, producing a set of concepts and categories simultaneously simpler and more powerful than the query transformation of classic information retrieval research, more explicit and discriminating than librarianship and indexing, particularly with regard to the significance and costs of human mental labor, and fuller, and more technologically informed, than information society discussions.

An inclusive understanding of information retrieval systems, developing from common understandings and conveyed by ostensive exemplification, rather than by restrictive definition, is adopted. In particular, the common antithesis between experimental and operational systems is dissolved. The real source of the contrast between the types of system would seem to have been the use of different forms of description process, particularly the experimental preference for machine generation rather than human selection of index terms and for non-Boolean operations on those descriptions in searching. Information retrieval systems and database management systems are also considered from a single perspective, with the commonality and contrast between them, the use of descriptions but with different schema for their construction, identified. Both sets of distinctions appear theoretically weak when made explicit and are increasingly eroded in practice, with operational systems possibly selecting records or documents by directly Boolean operations, but ordering retrieved documents on the basis of other indicators, and databases indexing attributes as well as entities.

Selection power

Selection power is understood as the human ability to make informed choices between objects or representations of objects. It is adopted here as the primary value or aim for information retrieval systems in contrast to the stress of query transformation in the experimental research tradition. The debate between the value of selection power and query transformation, which has been conducted elsewhere (Wilson, 1996c; Warner, 2000), is not pursued further here, other than to indicate that query transformation can be absorbed, both conceptually and operationally, into selection power as a special case within a more encompassing theory. Selection power is not arbitrarily asserted, its epistemological and ontological status is clarified, its content conveyed through exemplification, and its value supported by the presence of analogous concepts in separate scholarly and ordinary discourses.

Definition and elucidation

Selection power, like other fundamental concepts, may be difficult to define without becoming tautological. The difficulty of definition could be an intimation of the significance of the concept. Definition, in the classic sense of decomposition into more primitive concepts is deliberately avoided and would be difficult given the fundamental nature of the principle, but the term is still elucidated and a refusal of explication similarly avoided.

Questioning of the classic practice of definition as decomposing a concept into known entities can be found, somewhat playfully, in literary sources, partly by implication in linguistics, and, most explicitly, in philosophical texts. In *Great Expectations*, at the point where Pip is being inducted into written literacy, Dickens plays upon the inevitable circularity of definitions:

‘Your sister’s a master-mind. A master-mind.’

‘What’s that?’ I asked, in some hope of bringing him to a stand. But Joe was readier with his definition than I had expected, and completely stopped me by arguing circularly, and answering with a fixed look, ‘Her.’
 ‘And I ain’t a master-mind.’

(Dickens, 1946, p.50)

The association implied between the perception of a need for definitions and written literacy has been more formally stated, with an insistence on preliminary definitions regarded as a product of the cultural transition from orality to literacy (Goody and Watt, 1968).

A structuralist perspective in linguistics implies a finally circular understanding of meaning. For instance, Saussure in *The Course in General Linguistics* regarded linguistic signs as obtaining meaning from their negative differences from other signs, in a network of signs (Saussure, 1983; Culler, 1988, p.52). Most radically, and seemingly independently, Wittgenstein admitted the impossibility of defining primitive signs by further decomposition and advocated elucidation rather than definition:

The meaning of primitive signs can be explained by elucidations.
 Elucidations are propositions which contain the primitive signs. They can, therefore, only be understood when the meaning of these signs are already known.

(Wittgenstein, 1981, § 3.263)

Wittgenstein continues to pursue a rigorous logical development with regard to the possibilities of combination of primitive signs. The commitment here to a logical structure and formal, as well as discursive, presentation does not, then, imply an adherence to logical positivism under which empirical propositions must correspond to sense impressions (Ayer, 1936). Technically, elucidations rather than definitions will be offered.

The primitive signs to which Wittgenstein refers can be regarded as corresponding to atomic facts also distinguished in the *Tractatus Logico-Philosophicus*, and: ‘From the existence or non-existence of an atomic fact we cannot infer the existence or non-existence of another’ (Wittgenstein, 1981, § 2.062). Selection power is, then, to be received as an atomic fact or primitive proposition, not amenable to further decomposition, and elucidations rather than definitions are offered. A further implication of regarding selection power as an atomic fact would be that we cannot infer the existence of other atomic facts from it. The process of elucidation will begin with an example and then proceed to indicating concepts analogous to selection power in independent scholarly and ordinary discourses, implicitly acknowledging that practical understanding of how best to construct and use information systems has been in advance of theoretical articulation and striving to rise from the concrete to the abstract.

Example

An example can be given of the selection power which could be enabled by an information system. At one point in time, a researcher might wish to distinguish the private individual, Samuel Langhorne Clemens, from the author, Mark Twain. What would be valuable for this purpose would be a system which did not conflate these two distinguishable aspects of the individual but enabled them to be differentiated. At a later point in time, the same researcher might be interested in information on Mark Twain and Samuel Clemens considered as single entity. An information retrieval system should then be able to differentiate and to link together the occurrences of these different names, as required. The example was originally conceived as fictional in a double sense (Warner, 2000), but does have real historical roots. Twain's copyright disputes are indexed under his legal name of Clemens, although without a link to his pen name, in collections of copyright proceedings (Copyright Decisions, 1909). A generic search across different sources, seeking for Twain and Clemens as a single entity would then have to adapt the terminology adopted for searching to the particular source being used. Index terms can offer discriminations and links between related subjects, but selection power is still seen as a characteristic of human consciousness, derivable from, but not inhering in, semiotic products.

The relation of the nexus of Mark Twain and Samuel Clemens, the author and the private individual, to Mark Twain, exclusively the author, and Samuel Clemens, solely the private individual, corresponds to the genus: species relation, with public: not public as the differentiating factor. The relation could be represented diagrammatically (See Figure 1.). The genus species relation occurs repeatedly in indexing languages, for instance in thesaural relations between broader and narrower terms, and in the relation of indexing terms taken from a controlled vocabulary to the language of discourse, particularly as generic scope contrasted with specificity. For formal logic, the relation of species to genus is analogous to material implication ($p \text{ is a member of } q$ has similar truth conditions to $p \text{ implies } q$), although objects not classes could be denoted by the variables for material implication (Bell 1937, volume 2, p.491)). Material implication has itself been the most productive, although also the most difficult, of logical relations (Quine, 1937, p.84).

Discrimination between Twain the author and Clemens the private individual could be obtained by: a direct serial reading of relevant texts, with the searcher expending labor in reading and discrimination, possibly creating an index, in effect; by algorithmic and computer-conducted transformations on texts, with the searcher required to eliminate false recalls and to retrieve all relevant instances, both complicated by inevitable inconsistencies in the language of discourse; or by human assignment of index terms and references to sections of discourse, for interpretation by the searcher. At each stage distinguished, the selection power of the searcher is increased, their search labor reduced, and description labor and processes, either transferred to technology or embodied in a human indexer, are increased.

Different forms of graphic representation, pictorial, handwritten, or printed (See Figure 2.), would offer different possibilities for algorithmic transformation. Curiously, the

standard form of computer representation of written language, for instance, as ASCII code, which appears more finished and to retain less specific traces of its production than handwriting, is more amenable to algorithmic transformation as the process, for instance of keyboarding, is less congealed in the product, as storage in a form of computer memory. A further transformation, into more fully graphic rather than directly encoded representation, embodied in different file formats, represents a fuller congealing of the process in the product and would complicate algorithmic transformations.

A principle of selection power has, then, been shown to operate for an aspect of information retrieval practice often considered separately from other aspects, analogous to cataloging in distinction from classification or characterized as data contrasted with subject or topic retrieval. In addition to the value of a unified description, the recognition of the common principle cataloging shares with classification and subject determination restores theoretical significance and value to it, congruent with the often dominant actual use of information retrieval systems for identifying, recalling, and retrieving known items or the works of a given author (Smithson, 1994; Shneiderman, 2003, p.54).

Scholarly and ordinary discourses

Different, and partly independent, scholarly discourses have implicitly endorsed selection power as a design principle for information retrieval systems. The concept is also valued in ordinary discourse discussions of information retrieval.

The interconnected fields of librarianship and indexing have endorsed as central aims, in different ways and not necessarily fully explicitly, concepts analogous to or necessary components of selection power. Within librarianship, bibliographic control was seminally defined in the post-1945 period as ‘mastery over written and published records’ (Unesco/Library of Congress, 1950, p.1) and is strongly analogous to selection power. Without bibliography the ‘records of civilization would be an uncharted chaos of miscellaneous contributions to knowledge, unorganized and inapplicable to human needs’ (Unesco/Library of Congress, 1950, p. viii). At that stage of technological development, direct human intervention or labor was required for the creation of records and indexes for bibliographic organization. Unesco regarded itself as born into ‘appalling post-war bibliographic chaos’ (Murra, 1951, p.47) and the distribution of responsibility to national agencies, required to produce national bibliographies and allied works on a shared model, was conceived as the remedy for the chaos and as the path towards universal bibliographic control. A more recent, and less explicitly noticed, move towards universal bibliographic control, particularly of monographic literature than journal articles, in the growth of *WorldCat*, has been prompted more by internal dynamisms in the process, rather than by imposition, by the possibility of sharing the costs of human description labor, through distributing the products of that labor, as catalog records. Later sophisticated discussions, more directly concerned with selection power, also distinguished bibliographic control from bibliographic organization, with organization as the means to control (Wilson, 1968), reinforcing the sense of selection power as a property of human consciousnesses, enabled by, but not directly inhering in, organization imposed on data.

In discussions of indexing, index terms are valued for their discriminatory power and this could be considered analogous to the concern with differentia in classical logic.

Discriminatory power could also be regarded as an essential component or organizational factor enabling discriminatory control or selection power. The technological constraints of written and printed documents, as well as the need to reduce the cognitive labor of the searcher, compelled more concise index descriptions than the object described, with some exceptions, such as concordances.

Cybernetics, emerging its modern form in the immediate post-1945 period, concurrently with the formalization of bibliographic control and partly concerned with information technologies envisaged for enhancing bibliographic control, also emphasized control and navigation (Wiener, 1954). Cybernetics was understood to embrace the ‘complex of ideas’ represented by:

the study of language ... the study of messages as a means of controlling machinery and society, the development of computing machines, and other such automata, certain reflections upon psychology and the nervous system, and a tentative new theory of scientific method.

(Wiener, 1954, p.15)

The emphasis on control did not always separate human from machine discrimination, particularly in the subsequent development of cybernetics. The etymology of cybernetics, partly as given by Wiener, who coined the term in testimony to the Greek *kubernētēs* or ‘steersman’, understood primarily as a human control mechanism (Wiener, 1954, p.15), and, more strongly, through its link to Cybernesias, or the pilots’ festival held in honor of Theseus’ navigation to Athens (Warner, 2003, p.554), points to a deeper level of selection in collective human experience.

Further support for the significance of selection power can be derived from the etymology of intelligence. *Intelligence* can be traced to the Latin form, *inter-legere*, to choose from or between things (Stevens, 1998, p.66), strongly analogous to selection power and implying deliberate choice rather than domination by brute needs. Better known would be the account of the formation of the Roman military legion.

When the city was built, in the first place, Romulus divided all the multitude that were of age to bear arms into military companies, each company consisting of three thousand footmen and three hundred horsemen. Such a company was called a ‘legion’, because the warlike were *selected* out of all.

(Plutarch, 1914, p.123)

Division or differentiation of individuals is then a further characteristic of man in the *polis* in its initial realization as the city state. For the discussion here, intelligence is

conceived as a quality of human consciousness rather than inhering in the objects differentiated.

Contemporary ordinary discourse comments on information systems are highly significant but difficult to produce as evidence. Evaluative criteria may be given by implication rather than fully and directly articulated. Yet when a searcher complains that it is difficult to control the number of records retrieved, a principle of discriminatory power is being invoked. More explicitly, one spoken response to a presentation of the value of selection power was: 'that's the basis [an enhanced capacity for informed choice] on which people use systems anyway' (Warner, 2000). Extra-disciplinary written comments, not directly influenced by the query transformation tradition, can be regarded as embodying ordinary discourse concepts. For instance, a sociological study of communication in philosophy, which stresses the importance of direct oral communication between significant philosophers, criticizes literature discovery by 'indexing and abstracting services (whether in printed media or electronically on-line), which overload the channels rather than focusing them' (Collins, 1998, p.45). Ordinary discourse concepts, then, although elusive, support the value of selection power.

Studies within information science, particularly since the mid-1990s, have advocated selection, evaluation and filtering, rather than recall of all and only all the relevant documents, as appropriate aims for information retrieval (Wilson, 1996b; 1996c). The concept still needs to be fully operationalized (Griesbaum, 2000). The subsequent development of a labor theoretic approach to information retrieval can be regarded as an operationalization of selection power, in relation to real world considerations of human labor and the costs of that labor.

Summary

The value of selection power has been supported by indicating analogous concepts in partly separate scholarly discourses. Librarianship held a comparable concept in bibliographic control, indexing had endorsed the discriminatory power of a term, a crucial factor in obtaining control, cybernetics valued control and navigation. The value of selection power was supported by the etymology of intelligence, and, more specifically in relation to information retrieval systems, by ordinary discourse comments on information retrieval. The concept of selection power was elucidated and understood as the human faculty for discrimination. It is here endorsed as the primary design and evaluation principle for information systems. Query transformation could be absorbed within selection power, as a special case of its implementation, appropriate in certain circumstances and compelled by certain forms of technology, for instance, by the batch-processing methods of the 1950s.

The first proposition, then, for the development of a labor theoretic approach to information retrieval, is to assert the value of selection power, with selection power understood as a property of human consciousness and as a primitive proposition, open to elucidation but not to further decomposition.

Selection power and selection labor

The relation between selection power and selection labor is not conceived either ahistorically or independently of technology. For instance in primarily oral societies, forms of recitation, not equivalent to verbatim repetition, but still dependent on individual memory within a communal context, were crucial to the preservation of knowledge, by renewal (Goody and Watt, 1968). A transitional form, which both inherits elements from orality and anticipates characteristics of written literacy, can be found in the Icelandic Law-speaker. The Law-speaker was both required to recite the law and to answer queries on legal and parliamentary procedures, by oral pronouncements, on the basis of his memory of the law (Njal, 1960, pp.306-308). From a modern perspective, the Law-speaker could be regarded as an information system, embodied in an individual. With increased social complexity and the growth of documents, and of indexes to documents, direct mental labor in memory and recall is transferred to exosomatic sources. The knowledgeable person, variously conceived, may still remain significant to information seeking and offer a readiness and focus of response, difficult to obtain from more formalized information systems. Selection labor may be reemerging as a single category, concentrated in searcher labor, in a development concordant with other features of secondary orality (Ong, 1982).

The concern, then, is not directly with the mental labor of memory, recall, and response, although this is acknowledged as a form of labor, even if not subjectively experienced as such. Rather it is with the technological forms to which the cognitive burden of memory and recall has been transferred and with the mental labor involved in their construction and searching. In pre-modern practice (using the term *pre-modern* to refer to written and printed forms in distinction from computer-based or *modern* systems), physical labor in the organization of documentary materials was required, and, most significantly, there was a relatively clear distinction of description from search labor. The concern here is not with that physical labor, although, it is similarly acknowledged. The distinction between description and search labor will be carried forward as an analytical distinction, although the difficulty of substantive separation is recognized: for instance, a searcher requesting a list of documents in chronological order is effectively instantiating a form of description labor or process, at the point of searching. The contrast can be traced to the fixity of the technologies of writing, compared with the possible fluidity of computation (Warner, 2001, pp.33-46). The concern here will be with the theoretical possibilities for and constraints on selection labor, conceived as incorporating both description and search labor.

Theoretical minima for selection labor can be derived from serial possibilities. If items are examined serially, and without regress, then selection labor rises linearly with the number of objects in the collection examined. An unchanged principle for discrimination is assumed, rather than a conversational or dialogic alteration of the principle for discrimination, and this would be closely analogous to the batch processing historically practiced. An absence of meaningful organization of the objects discriminated and a conflation and simultaneous occurrence of description and search labor are also implied. If the choice between objects examined is reduced to a binary contrast between

acceptance and non-acceptance, then the unit of labor has some analogies with the classic understanding of the bit.

Imposing organization upon a collection of objects separates description from search labor, with work invested in organization or description reducing labor and enhancing power in searching. The overall quantity of labor is not altered but significantly redistributed. If objects can be partitioned by description and organization into appropriate sets, then Shannon's formula for the information of a source would indicate that the number of objects which could be discriminated by search labor would rise more than linearly with the quantity of search labor.

If there are N possibilities [for the choice of messages from a source], all equally likely, the amount of information is given by $\log_2 N$.

...

If it were possible to choose questions which always had the effect of subdividing into two equal groups, it would be possible to isolate, in twenty questions, one object from approximately 1,000,000 possibilities.

(Shannon, 1993b, pp.214-215)

On this basis, the number of choices, or, broadly, units of labor, required to discriminate between c.1000 and c.1,000,000 such possibilities, which could correspond to objects, documents, or records for documents, would double. The search could be conducted either deterministically or non-deterministically, with human intervention at intervals, both with unaltered criteria during the process. The labor invested in description corresponds to a capital cost, which not be incurred for each iteration of searching.

These slightly abstract considerations are helpful for establishing theoretical constraints for the labor associated with selection and for enforcing the point that labor can be distributed between description and searching, but cannot be done away with. They also have some more practical resonances. Semantic primitives have proved difficult to isolate, particularly for human or social discourse and may not even exist. In other aspects of information theory, for instance in reducing redundancy in messages, an approach to theoretical limits can be made, but they have proved difficult fully to obtain (Shannon, 1993a, p.39; Verdú and McLaughlin, 2000). The closest analogues to reduction to atomic facts, or to a perfectly organized source, might be offered by biological classifications although, even here, the assignment of objects to categories can be problematic (Darwin, 1968, pp.104-108). The non-linear relation between primitives distinguished for description and organization and the number of objects for description might be an underlying explanation for the possibilities of scaling offered by practical devices, such as thesauri, in working information systems.

The second proposition, then, for the development of a labor theoretic approach to information retrieval is that, with certain forms of exosomatic technologies, which tend to

be adopted under historical circumstances of increased social complexity, selection power is produced by selection labor. The production of selection power by selection labor, and the decomposition of selection labor into description and search labor, is particularly clearly exemplified within written literacy, but continues, in modified form, with modern information technologies.

Conclusion

Some fundamental issues have been addressed and propositions developed. Information retrieval research was reviewed and selection power endorsed as the primary value or aim for information retrieval. Selection power was received as a primitive proposition whose content was elucidated, and value supported, from analogous concepts in partly independent discourses and from its embodiment in information retrieval practice. Selection power was produced by selection labor. The concept of informational or mental labor has begun to be absorbed into an understanding of information retrieval. There is a congruence of values with processes, for information retrieval, particularly through the common idea of selection.

For the further development of a labor theoretic approach to information retrieval, the concept of selection labor needs to be further discriminated, with the distinction of description from search labor developed and exemplified, and the possibility of transfer of direct human labor to technology explored. Semantic mental labor will be distinguished from syntactic mental labor, with mental labor differentiated from the processes which can be abstracted from it and transferred to technology, and from the products which can be made from mental labor. These topics will be the concern of a subsequent article.

References

- Ayer, A.J. (1980). *Language, truth and logic*. Harmondsworth, England: Penguin Books.
- Bath University Library. (1971). *Information requirements of researchers in the social sciences*. Bath: Bath University Library.
- Bell, E.T. (1937). *Men of mathematics*. London etc.: Penguin Books.
- Buckland, M. and Plaunt, C. (1994). On the construction of selection systems. *Library Hi Tech*. 12, 15-28.
- Cleverdon, C. (1962). *Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems (An investigation supported by a grant to ASLIB by the National Science Foundation)*. Cranfield, College of Aeronautics.
- Cleverdon, C., Mills, J., Keen, M. (1966). *Factors determining the performance of indexing systems*. Cranfield, College of Aeronautics.
- Collins, R. (1998). *The sociology of philosophies: a global theory of intellectual change*. Cambridge, Massachusetts and London, England: The Belknap Press of Harvard University Press.
- Copyright Decisions. (1909). *Decisions of the United States Courts Involving Copyright and Literary Property 1789-1909*. Washington: Copyright Office, Library of Congress.
- Culler, J. (1988). *Saussure*. London: Fontana Press.
- Darwin, C. (1968). *The origin of species by means of natural selection or The preservation of favoured races in the struggle for life*. London etc.: Penguin Group.
- Dickens, C. (1946). *Great expectations*. London etc.: Cumberlege.
- Ellis, D. (1984). Theory and explanation in information retrieval research. *Journal of Information Science*. 8, 25-38.
- Ellis, D. (1989a). A behavioural model for information retrieval system design. *Journal of information science*. 15, 237-247.
- Ellis, D. (1989b). A behavioural approach to information retrieval system design. *Journal of documentation*. 45, 171-212.
- Ellis, D. (1992). The physical and cognitive paradigms in information retrieval research. *Journal of Documentation*. 48, 45-64.

- Ellis, D. (1996). Progress and problems in information retrieval. London: Library Association.
- Ellis, D. and Vasconcelos, A. (1999). Ranganathan and the Net: using facet analysis to search and organise the World Wide Web. *Aslib Proceedings*, 51, 3-10.
- Goody, J. and Watt, I. (1968). The consequences of literacy. In J. Goody (Ed.). *Literacy in traditional societies*. Cambridge: Cambridge University Press.
- Gosden, C. (2003). *Prehistory: a very short introduction*. Oxford: Oxford University Press.
- Griesbaum, J. (2000). Evaluierung hybrider Suchsysteme im WWW. iDiplomarbeit, Universität Konstanz Informationswissenschaft. Available at http://www.inf.uni-konstanz.de/~griesbau/files/evaluierung_hybrider_suchsysteme_im_www.pdf
- Heine, M.H. (1980). The 'question' as a fundamental variable in information science. In O. Harbo and L. Kajberg (Eds.), *Theory and application of information research (Proceedings of the Second International Research Forum on Information Science, 3-6 August 1977, Royal School of Librarianship, Copenhagen)* (pp.137-145). London: Mansell.
- Lyotard, J-F. (1984). *The postmodern condition: a report on knowledge*. Translated by G. Bennington and B. Massumi. Minneapolis: University of Minnesota Press.
- Minsky, M. (1967). *Computation: finite and infinite machines*. Englewood Cliffs, NJ: Prentice-Hall.
- Murra, K.O. (1951). History of some attempts to organize bibliography internationally. In J.H. Shera and M.E. Egan (Eds.). *Bibliographic organization: papers presented before the Fifteenth Annual Conference of the Graduate Library School July 24-29, 1950* (pp. 24-53). Chicago: University of Chicago Press.
- Njal. (1960). *Njal's Saga*. Translated and introduced by M. Magnusson and H. Pálsson. Harmondsworth etc.: Penguin Books.
- Ong, W.J. (1982). *Orality and literacy: the technologizing of the word*. London and New York: Methuen.
- Plutarch. (1914). *Plutarch's Lives*. English translation by B. Perrin. W. Heinemann and The Macmillan Company: London and New York. Volume 1.
- Quine, W.V.O. (1953). New foundations for mathematical logic. In W.V.O. Quine. *From a logical point of view: logico-philosophical essays*. Cambridge, Massachusetts: Harvard University Press.

- Railton, S. (2005). Mark Twain in his times. At <http://etext.lib.virginia.edu/railton/index2.html>, 17 March 2005.
- Ramsey, F.P. (1990). The foundations of mathematics (pp.164-224). In D. H. Mellor (Ed.). Philosophical papers. Cambridge University Press, Cambridge etc..
- Roberts, N. (1989). Online: in an educational cul-de-sac? Education for Information. 7, 101-106.
- Saussure, F. de (1983). Course in general linguistics. Edited by C. Bally and A. Sechehaye with the collaboration of A. Riedlinger, translated and annotated by R. Harris. London: Duckworth.
- Shannon, C.E. (1993a). A mathematical theory of communication. First published 1948. In C.E. Shannon, Collected papers (pp.5-83). Edited by N.J.A. Sloane & A.D. Wyner. Piscataway, NJ: IEEE Press.
- Shannon, C.E. (1993b). Information theory. In C.E. Shannon. Collected papers (pp.212-220). Edited by N.J.A. Sloane and Aaron D. Wyner. Piscataway, NJ: IEEE Press.
- Shneiderman, B. (2003). Leonardo's laptop: human needs and the new computing technologies. Cambridge, Massachusetts and London, England: MIT Press.
- Smithson, S. (1994). Information retrieval evaluation in practice: a case study approach. Information Processing and Management. 30, 205-221.
- Stevens, A. (1998). Ariadne's clue: a guide to the symbols of mankind. London etc.: Allen Lane, The Penguin Press.
- Unesco/Library of Congress. (1950). Bibliographical services, their present state and possibilities of improvement: report prepared as working paper for an international conference on bibliography. Washington, DC.
- Verdú, S. and McLaughlin, S.W. (2000). Information theory: 50 years of discovery. New York: IEEE Press.
- Vico, G. (1988). On the most ancient wisdom of the Italians: Unearthed from the origins of the Latin language: including the disputation with the Giornale de' Letterati d'Italia. First published 1710. Translated and introduced by L.M. Palmer. Ithaca and London: Cornell University Press.
- Warner, J. (1994). From writing to computers. London and New York: Routledge.
- Warner, J. (2000). In the catalogue ye go for men: evaluation criteria for information retrieval systems. Aslib Proceedings. 52, 76-82.

- Warner, J. (2001). Information, knowledge, text. Lanham, MD: Scarecrow Press.
- Warner, J. (2003). Information and redundancy in the legend of Theseus. *Journal of Documentation*. 59, 540-557.
- Warner, J. (2004). Humanizing information technology. Lanham, MD: Scarecrow Press.
- Warner, J. (2005a). Labor in information systems. *Annual Review of Information Science and Technology*. 39, 551-573.
- Warner, J. (2005b). An information dynamic: technologies for the reproduction of written utterances. *Aslib Proceedings*. 57, 2005, 5, pp.412-423.
- Webster, F. (2002). Theories of the information society. 2nd edition. London and New York: Routledge.
- Wiener, N. (1954). The human use of human beings: cybernetics and society. Revised edition. New York: De Capo Press.
- Wilson, P. (1968). Two kinds of power: an essay on bibliographical control. Berkeley and Los Angeles: University of California Press.
- Wilson, P. (1973). Situational relevance. *Information storage and retrieval*. 9, 457-471.
- Wilson, P. (1996a). Review of F. Webster. Theories of the information society. *College and Research Libraries*. 57, 487-489.
- Wilson, P. (1996b). Interdisciplinary research and information overload. *Library Trends*. 45, 192-203.
- Wilson, P. (1996c). Some consequences of information overload and rapid conceptual change. In J. Olaisen, E. Munch-Petersen, and P. Wilson (Eds.). *Information science: from the development of the discipline to social interaction* (pp.21-34). Oslo etc.: Scandinavian University Press.
- Wilson, P. (2001). Review of E. Svenonius. The intellectual foundations of information organization. *College and Research Libraries*, 62, 203-204.
- Wittgenstein, L. (1981). *Tractatus Logico-Philosophicus*. London and New York: Routledge and Kegan Paul.

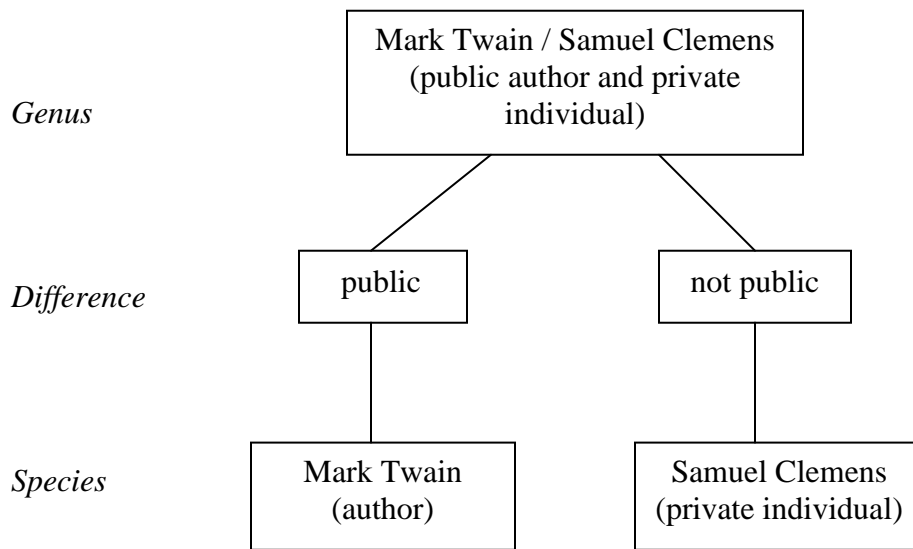


Figure 1. Example of a genus-species relation.

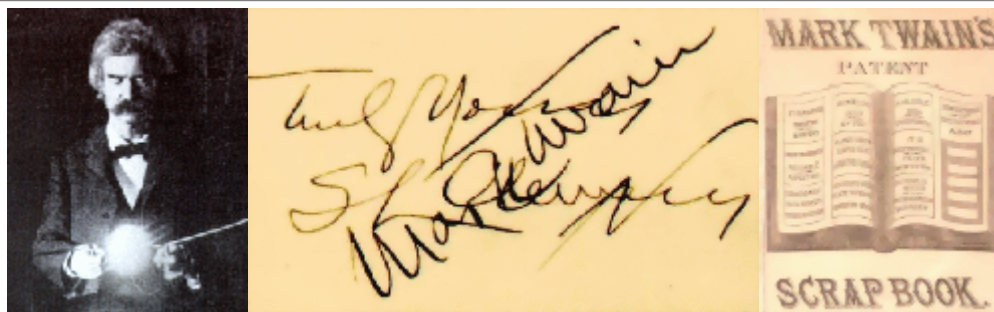


Figure 2. Contrasting representations of Samuel Clemens / Mark Twain (Source: Railton, 2005).

Acknowledgments

I would like to acknowledge the John Campbell Trust who supported a presentation of the themes of this article at the Annual Meeting of the American Society for Information Science and Technology Annual Meeting, Providence, Rhode Island, November 2004, and the hospitality and assistance of the Research Centre for the Social Sciences (RCSS), University of Edinburgh, where the article was originally written during study leave from the Queen's University of Belfast, February–July 2005.